

# Прогнозування кредитоспроможності клієнтів банків за СКОРИНГОВИМ МЕТОДОМ

*Виконав: Ревва Роман Володимирович, група КА-34*

*Науковий керівник: д.т.н., проф., Бідюк Петро Іванович.*



# Актуальність дослідження

- ▶ Кредитний ринок в Україні є основним сегментом фінансового ринку, що передбачає можливість для суб'єктів господарювання швидко мобілізувати фінансові ресурси. Наявність розвинутого та ефективно функціонуючого кредитного ринку є основою активізації підприємницької діяльності та економічного розвитку держави в цілому.
- ▶ Банки виступають як у ролі покупця, так і ролі продавця на кредитному ринку. Кількість банків у банківській системі України за період 2012-2016 рр. зменшилась майже удвічі. Станом на 01 січня 2017 року ліцензію Національного банку України мали 96 банківських установ (в т.ч. 38 банків з іноземним капіталом). З початку 2016 року кількість функціонуючих банківських установ скоротилася на 21. Загалом, з початку 2014 року внаслідок погіршення платоспроможності до 82 банківських установ було запроваджено тимчасову адміністрацію.



# Актуальність задачі

- ▶ Велика кількість неповернених коштів через дефолт позичальників кредитів;
- ▶ Недостатня автоматизація процесу обробки заявок на отримання кредиту в режимі реального часу;
- ▶ Високий рівень складності існуючих систем кредитного скорингу у режимі реального часу для кінцевого користувача;

# Мета і об'єкт дослідження

- **Мета:** розробка системи оцінювання та прогнозування кредитоспроможності клієнтів у режимі реального часу використовуючи скоринговий метод
- **Предмет:** Математичні моделі, методи, і критерії оцінювання адекватності скорингових моделей та методи побудови скорингової карти
- **Об'єкт:** Позичальники кредитів, представлені статистичними даними з вибраними характеристиками



# Задачі

- 1. Вибір алгоритму підготовки даних**
- 2. Визначення критеріїв оцінювання точності моделей**
- 3. Оцінювання параметрів моделей та вибір найбільш точної скорингової моделі**
- 4. Розробка автоматизованої системи прогнозування кредитоспроможності клієнтів банку**



# Вибір моделей

## ► Досліджувані моделі дослідження:

1. Логістична регресія
2. Пробіт-регресія
3. Нейронні мережі
4. Метод найближчих к-середніх
5. Метод дерев рішень

# Логістична регресія: логіт-модель

Загальний вигляд моделі:

$$\mathbb{P}\{y = 1 \mid x\} = f(z)$$

Логіт-функція

$$f(x) = \frac{1}{1 + e^{-x}}$$

Залежність вихідної змінної від значень параметрів моделі

$$z = \theta^T x$$



# Логістична регресія: пробіт-модель

Загальний вигляд моделі:

$$\mathbb{P}\{y = 1 \mid x\} = f(z)$$

Пробіт-функція

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

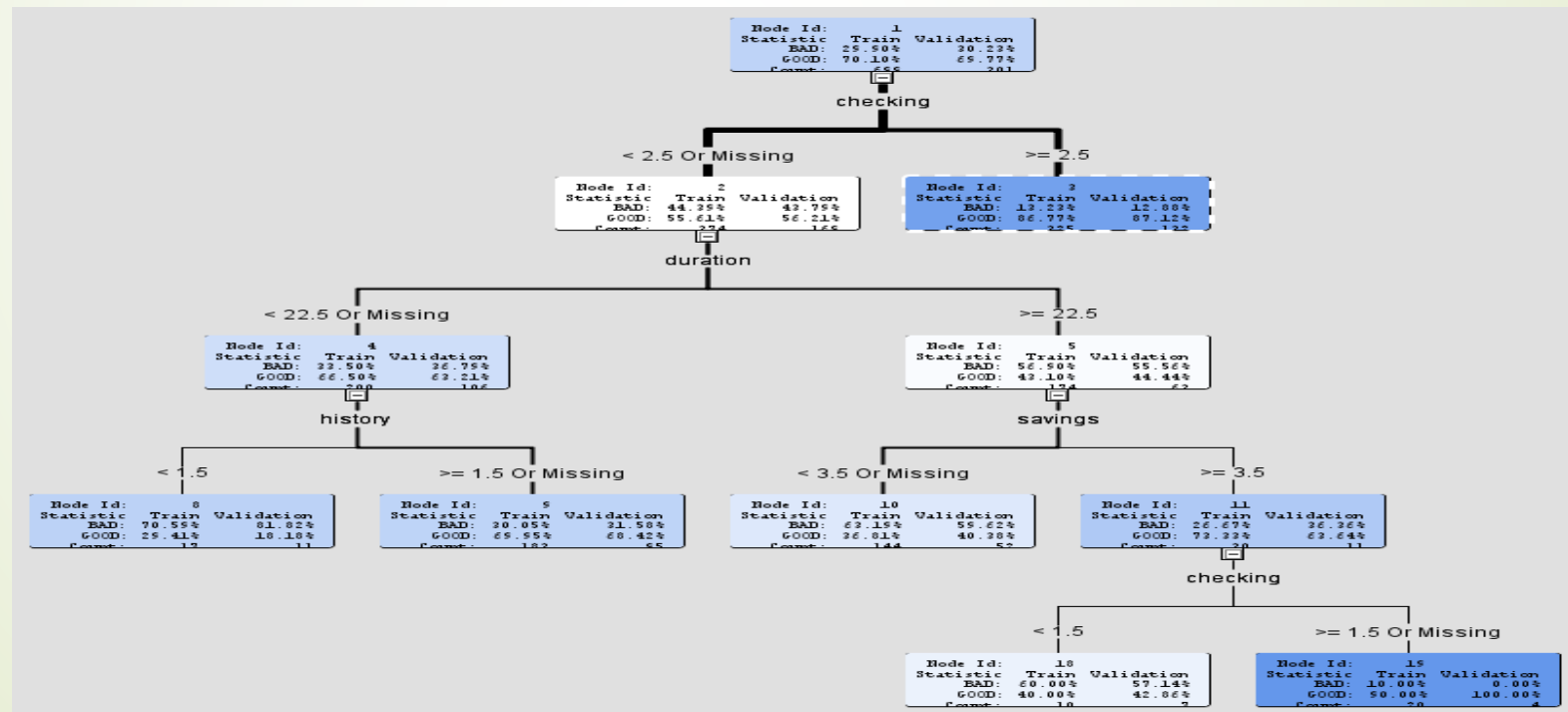
Залежність вихідної змінної від значень параметрів моделі

$$z = \theta^T x$$



# Дерева рішень

- Дерево рішень – це модель, що будується на логічному ланцюжку правил, які призначені для опису окремих взаємозв'язків між даними щодо очікуваного результату. Графічно спрощена модель показана на рисунку.



# Критерії оцінки моделей

- **Misclassification Rate (MR)** – загальна точність моделі, що визначається як відношення кількості хибно спрогнозованих значень цільової змінної до загальної кількості елементів вибірки;
- **True Positive Rate (TPR)** – відсоток вірно спрогнозованих позитивних значень цільової змінної;
- **True Negative Rate (TNR)** – відсоток вірно спрогнозованих негативних значень цільової змінної;

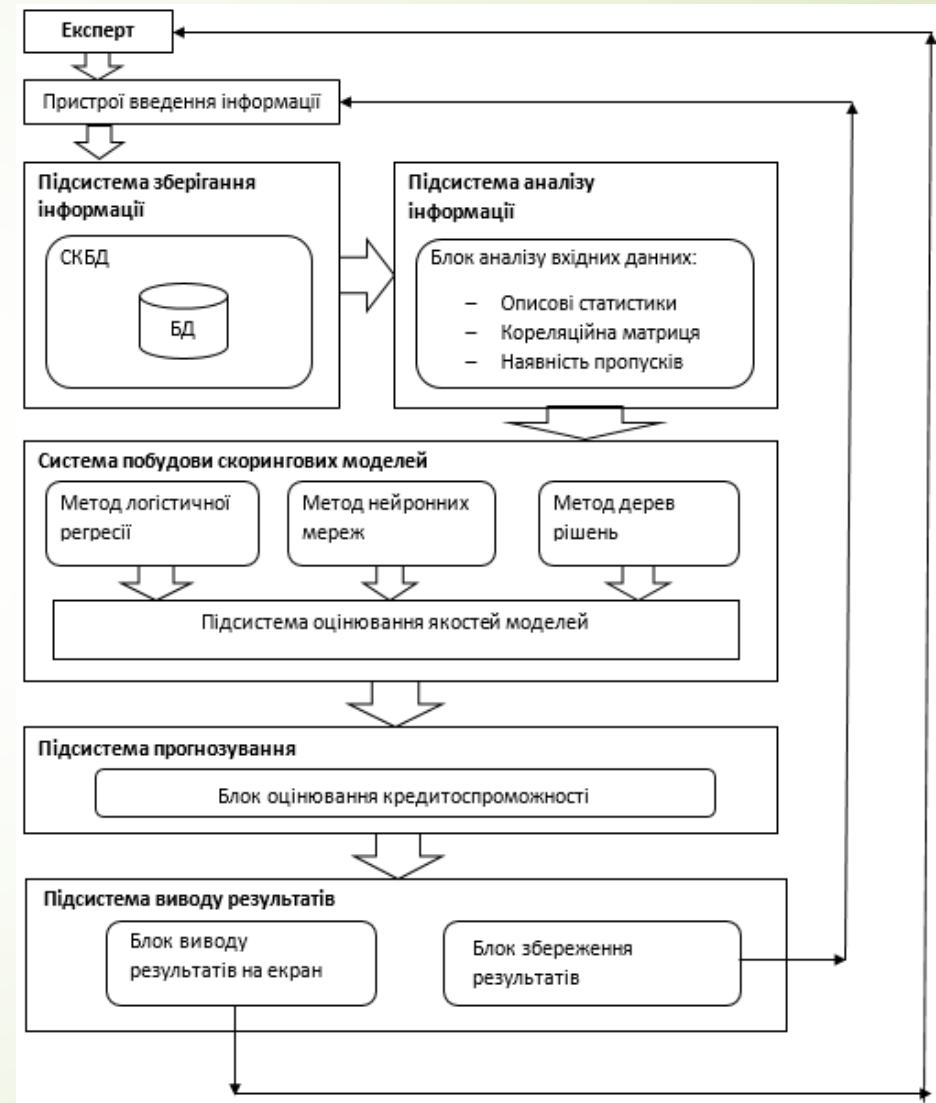


# Історична вибірка вхідних даних

- ▶ **Вибірка історичних анонімізованих соціо-демографічних даних клієнтів одного з українських банків;**
- ▶ **Об'єм вибірки:** 1000 записів;
- ▶ **Кількість змінних:** 21;
- ▶ **Формат вхідного файлу даних:** .csv;
- ▶ **Цільова змінна:** good\_bad (індикатор дефолту клієнта);

# Архітектура розробленого продукту

- Система є складною та складається з чотирьох рівнів:
- Підсистема зберігання вхідних даних;
- Підсистема аналізу вхідних даних;
- Система побудови скорингових моделей;
- Система прогнозування та виводу результатів



# Интерфейс розробленого продукту

Дипломна Робота Ревва КА-34

Файл Справка

### Исходные данные

	Checking	Duration	History	Purpose	Amount
▶	1	6	4	3	1169
	2	48	2	3	5951
	4	12	4	6	2096
	1	42	2	2	7882
	1	24	3	0	4870
	4	36	2	6	9055
	4	24	2	2	2835
	2	36	2	1	6948
	4	12	2	3	3059

### Корреляционная матрица:

Checking	Duration	History	Purpose	Amount
1	0,756553869523...	0,856111474294...	0,654093124866...	0,667066231215
0,756553869523...	1	0,777182448837...	0,658079246609...	0,858246917397
0,856111474294...	0,777182448837...	1	0,629398255271...	0,681492101316
0,654093124866...	0,658079246609...	0,629398255271...	1	0,544842977595
0,667066231215...	0,858246917397...	0,681492101316...	0,544842977595...	1
0,772269692072...	0,698353061159...	0,739344076485...	0,559882850608...	0,631642468770
0,859444840933...	0,821730123976...	0,882695935137...	0,676633262893...	0,712294246823
0,835564822939...	0,819334865366...	0,863097828681...	0,680392643379...	0,643376086601
0,869960745189...	0,834572244744...	0,890071156734...	0,688939530729...	0,729492963854

Посчитать характеристики

Введите данные о клиенте

	Checking	Duration	History	Purpose	Amount	Savings	Employed	InstallP	Marital	Coapp	Resident
▶											

Оценить кредитоспособность

### Результат

Прогноз:  
Вероятность дефолта:

# Интерфейс розробленого продукту

Дипломна Робота Ревва КА-34

Файл Справка

### Исходные данные

	Checking	Duration	History	Purpose	Amount
▶	1	6	4	3	1169
	2	48	2	3	5951
	4	12	4	6	2096
	1	42	2	2	7882
	1	24	3	0	4870
	4	36	2	6	9055
	4	24	2	2	2835
	2	36	2	1	6948
	4	12	2	3	3059

### Корреляционная матрица:

Checking	Duration	History	Purpose	Amount
1	0,756553869523...	0,856111474294...	0,654093124866...	0,667066231215
0,756553869523...	1	0,777182448837...	0,658079246609...	0,858246917397
0,856111474294...	0,777182448837...	1	0,629398255271...	0,681492101316
0,654093124866...	0,658079246609...	0,629398255271...	1	0,544842977595
0,667066231215...	0,858246917397...	0,681492101316...	0,544842977595...	1
0,772269692072...	0,698353061159...	0,739344076485...	0,559882850608...	0,631642468770
0,859444840933...	0,821730123976...	0,882695935137...	0,676633262893...	0,712294246823
0,835564822939...	0,819334865366...	0,863097828681...	0,680392643379...	0,643376086601
0,869960745189...	0,834572244744...	0,890071156734...	0,688939530729...	0,729492963854

Посчитать характеристики

### Введите данные о клиенте

	Checking	Duration	History	Purpose	Amount	Savings	Employed	InstallP	Marital	Coapp	Resident
▶	1	20	0	3	200000	10000	0	1	0	0	1

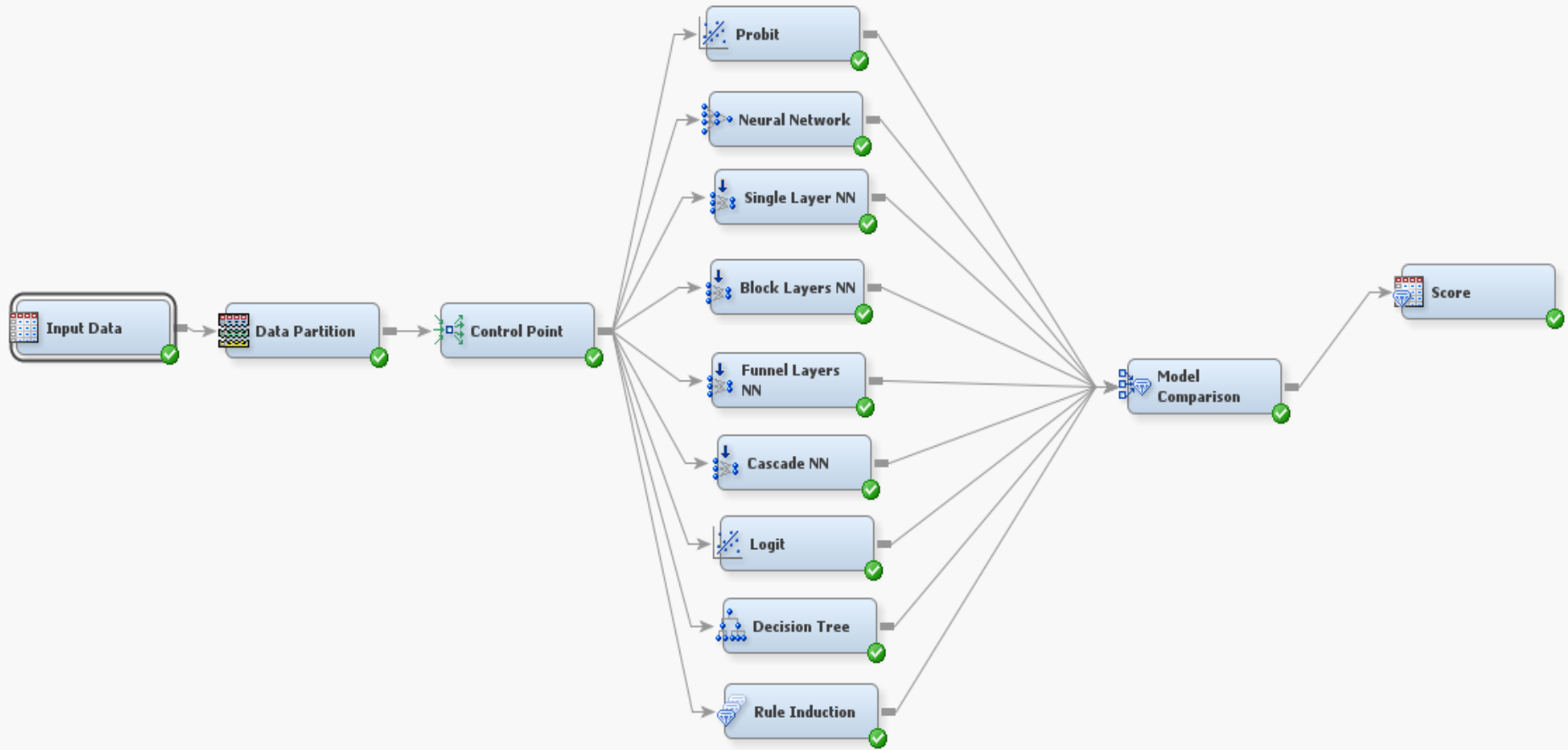
Оценить кредитоспособность

### Результат

Прогноз: **Клиент неплатежеспособен. Риск дефолта критический!**

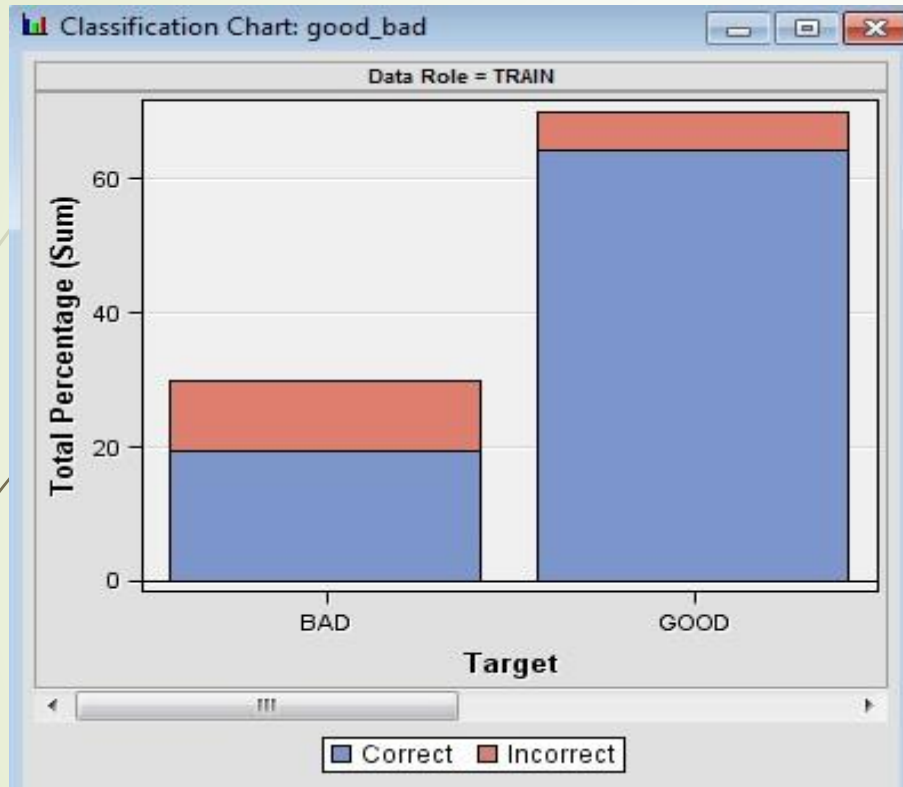
Вероятность дефолта: *Риск дефолта - 47%*

# Розглянуті моделі





# Результати: Логіт-регресія

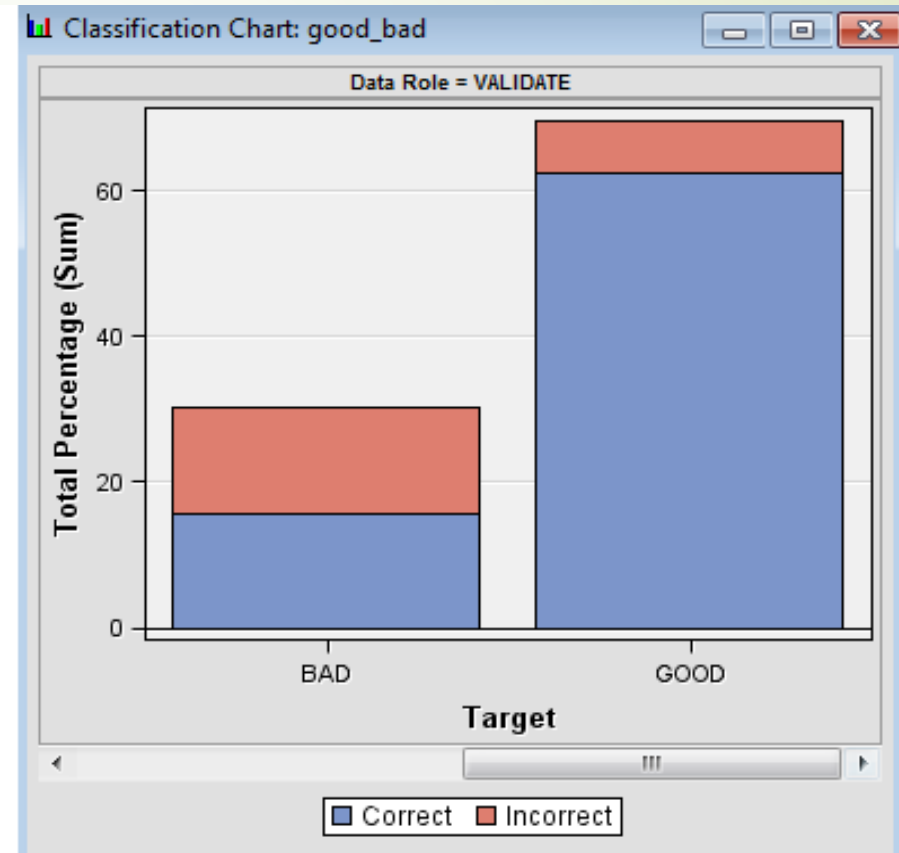
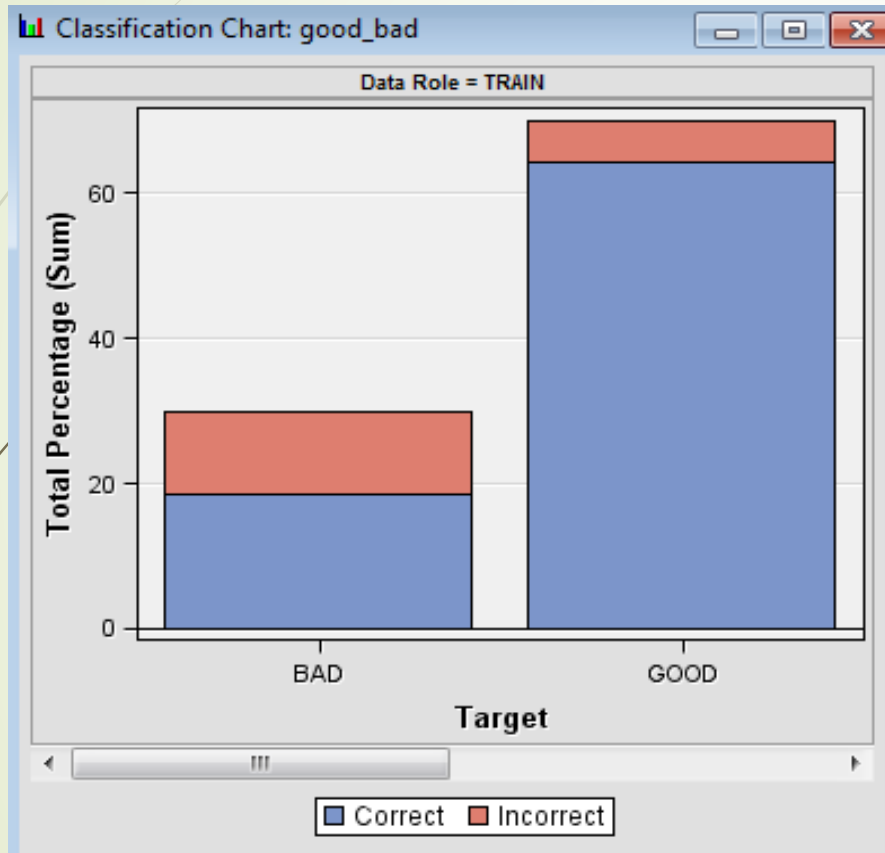


**MR:** 0,1645

**TNR:** 66%

**TPR:** 91%

# Результати: Пробіт-регресія



**MR:** 0,1745

**TNR:** 63%

**TPR:** 89%

# Результати: Дерево рішень



**MR: 0,23176**



**TNR: 54%**

**TPR: 83%**

# Результати порівняння моделей

Selected Model	Predecessor Node	Model Node	Model Description	Target Variable	Target Label	Selection Criterion: Valid: Misclassification Rate	Train: Akaike's Information Criterion	Train: Average Squared Error	Train: Average Error Function
Y	Reg2	Reg2	Logit	good_bad		0.164521	575.6057	0.116718	0.367386
	Reg	Reg	Probit	good_bad		0.174535	601.9616	0.126159	0.391961
	Tree	Tree	Decision Tr...	good_bad		0.23176	.	0.167387	.
	Rule	Rule	Rule Inducti...	good_bad		0.23176	.	0.158111	.
	AutoNeural2	AutoNeural2	Block Layer...	good_bad		0.238913	707.2074	0.153955	0.464383
	AutoNeural3	AutoNeural3	Funnel Lay...	good_bad		0.238913	707.2074	0.153955	0.464383
	AutoNeural	AutoNeural	Neural Net...	good_bad		0.286123	1210.077	0.218319	0.77831
	AutoNeural4	AutoNeural4	Cascade NN	good_bad		0.321888	1187.593	0.242096	0.760796

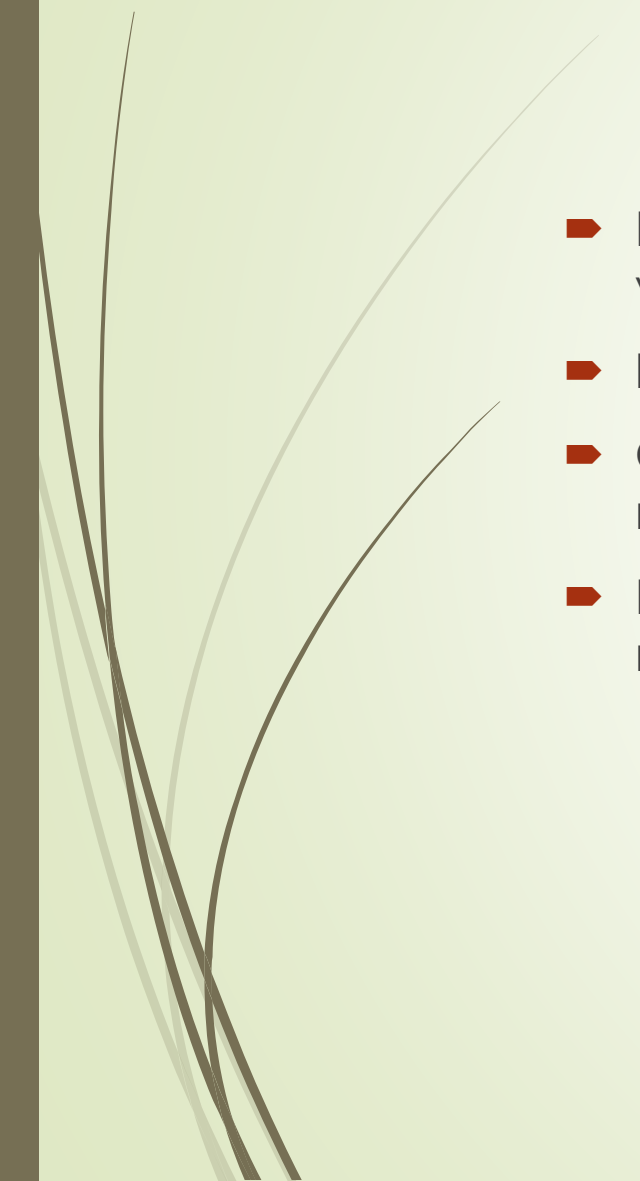



# Висновки

- ▶ Виконано аналіз функціонування кредитування у банківській системі України та встановлено актуальність і **необхідність** створення зручної для користувача скорингової системи в режимі реального часу;
- ▶ Проведено аналіз найбільш поширених скорингових моделей, побудовано кращу модель логістичної регресії з точністю **84%**;
- ▶ Розроблено швидкодіючу систему оцінювання та прогнозування кредитоспроможності клієнтів у режимі реального часу використовуючи скоринговий метод та середовище **.NET**;
- ▶ Швидкість аналізу кредитоспроможності клієнтів становить **1,34с**, що робить її максимально комфортною для роботи у режимі реального часу;



# Напрями подальших досліджень

- ▶ Підвищення точності існуючої моделі логістичної регресії шляхом уточнення її структури і параметрів;
  - ▶ Інтеграція системи з системами керування базами даних;
  - ▶ Сентиментальний аналіз записів клієнта у соціальних мережах, використовуючи засоби текстової аналітики для коригування прогнозу;
  - ▶ Географічний аналіз записів клієнта у соціальних мережах для коригування прогнозу;
- 



# Публікації та впровадження результатів дослідження

- ▶ Результати досліджень були частково використані у роботі **«Подготовка интерактивных отчётов для последующего упрощения процесса принятия решений в бизнесе»** в рамках виступу на конференції **SAIT-2017**
- ▶ Акт впровадження та використання результатів дослідження у проектах компанії **SAS Institute LLC**





Дякую за увагу!

Ревва Роман Володимирович, КА-34